On factor-augmented univariate forecasting

– International Panel Data Conference 2014 –

A-A. NIANG D. LEGROS M-C. PICHERY

Tokyo, July 9 - 10

SUMMARY

1. Motivation
2. Framework
3. Methodology
4. Forecast Accuracy
5. Monte Carlo Study
6. Concluding remarks

Common factor model

- Consider y_t, a vector of n stochastic variables with high degrees of cross-section dependences. Its dynamics can be represented as

$$y_{it} = \mu_i + \gamma_i f_t + z_{it}. \qquad (t = 1, ..., T, \quad i = 1, ..., n) \qquad (1)$$

  - f_t referred to as an unobserved common factor
  - μ_j is an individual fixed effect
  - z_{jt} idiosyncratic factor
  - γ_j, factor loading which give a measure of the contribution of the j-th individual to the common shock

- This is an attractive representation which allows :
  - to model large dimensional dataset with high degree of cross-sectional correlations
  - a parsimonious econometric structure
- An important upsurge of forecasting methods using common factors is noted in recent years : Stock and Watson (2002), Peña and Poncela (2004)...

Common factors as predictors

- A popular case is the diffusion index forecast (Stock and Watson, 2002). Consider the case with a single factor

$$y_t = \alpha + \beta f_{t-1} + \rho y_{t-1} + \xi_t \tag{2}$$

  - $|\rho| < 1$,
  - f_t is extracted from x_t a large panel of stationary time series which admits a common factor structure (Equation 1).

- The corresponding h-step ahead forecast is

$$\hat{y}_{t+h} = \alpha_h + \beta_h f_t + \rho_h y_t \tag{3}$$

  - β_h and ρ_h depend on the forecast horizon
  - $E\left(\xi_{t+h} | \{f_\tau, y_\tau\}_{\tau \leq t}\right) = 0$.

- Important forecast performance shown by the literature (improving forecast accuracy, outperforming many competing methods)

In the paper

- Augment univariate series by idiosyncratic factor : Idiosyncratic Factor-Augmented (IFA) model.
- Thus, an approach which focuses on the role of idiosyncratic factor.
- Show that IFA model also should work well for forecasting and represents an improvement with respect to the simple univariate model, since it permits to model country specific factors which should play an important role in the evolution of the univariate macroeconomic time series such as GNP.

Setup

- Consider that n in model (1) is enough large and that y_t is a vector of macroeconomic variables

$$y_{it} = \mu_i + \gamma_i f_t + z_{it}.$$

- The factor is allowed to follow a dynamic stationary vector process

$$f_t = \varphi f_{t-1} + \eta_t \qquad (4)$$

- Using these two last equations

$$y_{it} = (1-\rho)\,\mu_i + \gamma_i\,(\varphi - \rho)\, f_{t-1} + \rho y_{i,t-1} + z_{it} - \rho z_{i,t-1} + \gamma_i \eta_t. \qquad (5)$$

- univariate regression is a special case of the factor augmented model with the restriction $\varphi = \rho$ and $\gamma_i = 0$.
- diffusion index forecast, case where f_t and y_{it} are not restricted to have the same order of integration and where the restriction that η_t and $(1-\rho L)\, z_{it}$ are unpredictable is considered.

Forecasting model

- Single factor residual model (set $\varphi = \rho$)

$$\begin{aligned} y_{it} &= \alpha_i + \rho y_{i,t-1} + \xi_{it} \\ \xi_{it} &= \gamma_i \eta_t + e_{it} \end{aligned} \tag{6}$$

  - $\alpha_i = (1-\rho)\,\mu_i$
  - e_{it} follows MA process : $e_{it} = z_{it} - \rho z_{i,t-1}$.
- The residual is exploited for forecasting purpose

$$\hat{y}_{i,t+1} = \alpha_i + \rho y_{it} + \phi z_{it} \qquad \text{(one-step ahead)}$$

$$\hat{y}_{i,t+h} = \sum_{j=0}^{h-1} \rho^j \alpha_i + \rho^h y_{it} + \rho^{h-1} \phi z_{it}. \qquad \text{(h-step ahead)}$$

- With prediction mean square error

$$\mathcal{MSFE}_i^{(1)} = \sigma_{zi}^2 + \sum_{j=0}^{h-1} \rho^{2j} \frac{(\rho-\delta_i)(1-\rho\delta_i)}{\delta_i} \sigma_{zi}^2. \tag{7}$$

Univariate forecast vs. IFA forecast

- If the effect of the factor residual is mistakenly ignored , then the resulting univariate series will follows an ARMA (1,1) process

$$y_{it} = c_i + \rho y_{i,t-1} + v_{it} - \delta_i v_{i,t-1} \tag{8}$$

  - $|\delta_i| < 1$
  - $v_{it} \sim i.i.d. \left(0, \sigma_{vi}^2\right) \; \forall i$

- Univariate ARMA(1,1) forecast

$$\hat{y}_{i,t+1} = c_i + \rho y_{it} - \delta_i v_{it} \qquad \text{(one-step ahead)}$$

$$\hat{y}_{i,t+h} = \sum_{j=0}^{h-1} \rho^j c_i + \rho^h y_{it} - \rho^{h-1} \delta_i v_{it} \qquad (h\text{-step ahead})$$

- With prediction mean square error

$$\mathcal{MSFE}_i^{(2)} = \frac{(2\rho - \delta_i)}{\rho} \sigma_{zi}^2 + \sum_{j=0}^{h-1} \rho^{2j} \frac{(\rho - \delta_i)^2}{\rho \delta_i} \sigma_{zi}^2. \tag{9}$$

Univariate ARMA forecast vs. IFA forecast

- Let $\Delta_i = \mathcal{MSFE}_i^{(2)} - \mathcal{MSFE}_i^{(1)}$ be the measure of the forecast performance of the augmented univariate model, with respect to the simple univariate model. For the h-steps ahead forecast, we have :

  - In the non stationary case ($\rho = 1$),

$$\Delta_i = (1 - \delta_i)\,\sigma_{zi}^2 > 0. \tag{10}$$

  - In the stationary case ($|\rho| < 1$),

$$\Delta_i = \rho^{2h-1}\,(\rho - \delta_i)\,\sigma_{zi}^2 > 0. \tag{11}$$

- Reduction of the Mean Square Forecast Error
- When the horizon of prediction approaches infinity, the difference between both models vanishes (In the stationary case).
- A larger value of σ_{zi}^2 implies a greater forecast precision of the IFA model

Simulations design

- Three models are compared :

  (1) naive forecast

$$\hat{y}_{i,t+1} = \alpha_i^{AR} + \rho^{AR} y_{i,t},$$

  (2) ARMA(1,1) forecast

$$\hat{y}_{i,t+1} = \alpha_i^{ARMA} + \rho^{ARMA} y_{i,t} - \delta_i v_{i,t},$$

  (3) IFA forecast

$$\hat{y}_{i,t+1} = \alpha_i^{IFA} + \rho^{IFA} y_{i,t} + \phi z_{it}.$$

- 1,000 replications

Simulations results

Model	AR(1)			ARMA(1,1)			IFA		
ρ/h	1	5	10	1	5	10	1	5	10
				$(n=10)$					
0.2	1.3460	1.4054	1.4200	1.3499	1.4060	1.4203	1.3416	1.4046	1.4196
0.4	1.3857	1.4604	1.4695	1.3849	1.4598	1.4691	1.3736	1.4579	1.4682
0.6	1.3906	1.5074	1.5506	1.3792	1.5047	1.5493	1.3494	1.4985	1.5464
0.8	1.4863	1.7096	1.8050	1.4492	1.6996	1.8004	1.3981	1.6849	1.7934
1.0	1.6721	2.1428	2.6389	1.5268	2.0397	2.5682	3.0255	3.3633	3.7441
				$(n=20)$					
0.2	1.3307	1.4038	1.4166	1.3363	1.4048	1.4171	1.3263	1.4030	1.4162
0.4	1.3608	1.4428	1.4603	1.3566	1.4417	1.4597	1.3447	1.4398	1.4588
0.6	1.4359	1.5461	1.5841	1.4190	1.5417	1.5820	1.3914	1.5369	1.5797
0.8	1.4846	1.7197	1.8215	1.4454	1.7100	1.8171	1.3847	1.6941	1.8094
1.0	1.6753	2.1609	2.6761	1.5213	2.0461	2.5944	3.1742	3.5183	3.8931

Simulations results

$(n = 50)$

0.2	1.3722	1.4134	1.4315	1.3762	1.4141	1.4318	1.3681	1.4127	1.4311
0.4	1.3963	1.4542	1.4811	1.3936	1.4533	1.4806	1.3799	1.4510	1.4795
0.6	1.4434	1.5409	1.5895	1.4280	1.5368	1.5875	1.3986	1.5313	1.5849
0.8	1.5297	1.7293	1.8486	1.4884	1.7182	1.8434	1.4250	1.7014	1.8355
1.0	1.7034	2.2056	2.7643	1.5642	2.0931	2.6797	3.3656	3.6656	4.0529

Notes: The values in the table correspond to the square root of $\widehat{\mathcal{MSFE}}$ for AR(1), ARMA(1,1) and IFA models computed using equation (20). The number of Monte Carlo repetitions is 1,000 and in each draw the data generated are used to estimate the three models.

Simulations results

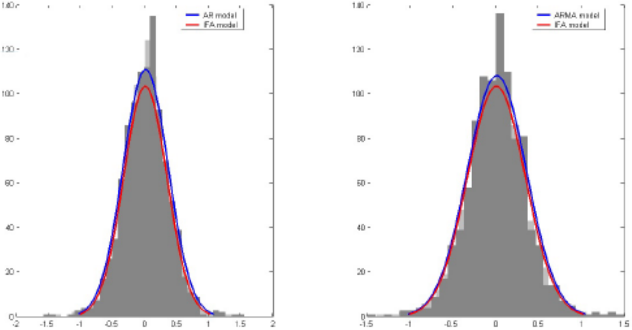

Figure 1: Distribution fit of forecast errors

Main findings

- Relative advantage of the IFA over the ARMA and the AR due to omission of the residual common factor structure.
- IFA model yields more accurate prediction both for near zero and near unit values of ρ.
- AR never the best model

Concluding Remarks

- A gain in precision, in terms of the prediction MSFE, of the IFA model with respect to ARMA and simple univariate depends on the importance of the share of variance of the idiosyncratic element.
- In the case of nonstationary series, results are very mitigated du to estimations problems linked with the presence of unit root.
- In all cases, AR never the best model